EXHIBIT 8

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MASSACHUSETTS

In re: Credit Suisse-AOL Securities Litigation	)	Civ. Action No. 02-12146-NG (Judge Gertner)

Rebuttal Declaration of M. Laurentius Marais, Ph.D.

I, M. Laurentius Marais, declare and state as follows:

1. I am a California resident over the age of 18. I have personal knowledge of the matters set forth in this declaration. I could competently testify regarding these matters if called to do so.

2. I am a Vice President and Principal Consultant at William E. Wecker Associates, Inc., a consulting firm specializing in applied mathematical and statistical analysis. I hold a Ph.D. degree and master's degrees in Business Administration, Mathematics, and Statistics from Stanford University. I have taught and conducted scholarly research while serving on the faculties of the University of Chicago and Stanford University. The subjects of my teaching and research have included the design and conduct of "event studies" of the effects of information on security prices. My qualifications and a list of my professional publications are shown in my curriculum vitae, which is attached to this declaration as Exhibit A.

3. I was asked by counsel for the plaintiffs in this matter to do the following:

(a) Review Dr. Scott M. Hakala's description in his March 4, 2008, Expert Report in this case of an "event study" that he performed in connection with his analysis of damages related to AOL's common shares. Dr. Hakala describes his event study in ¶ 29-35 on pp. 27-33 of his Expert Report.

(b) Review the portion of Dr. René M. Stulz's May 1, 2008, Expert Report in this case in which Dr. Stulz criticizes Dr. Hakala's use in his event study of "dummy variables" representing the dates of a set of "material events" concerning AOL, and opines that Dr. Hakala's event study results are "biased" and "unreliable" because of this aspect of its design. Dr. Stulz's criticism of this aspect of the Hakala event study appears in ¶ 99-101 on pp. 49-50 of his Expert Report.

(c) In light of (a) and (b), assess the validity of Dr. Stulz's criticism as a basis for his opinion that the Hakala event study results are "biased" and "unreliable."

4. In connection with this assignment, I reviewed the documents listed in Exhibit B.

Summary of Opinions

5. Dr. Stulz claims that Dr. Hakala's use of multiple "dummy variables" to guard against the potential, power-reducing effects of unrelated material events is flawed in that this procedure necessarily "biases" the event study results. Dr. Stulz's claim is incorrect.

6. Dr. Stulz claims that Dr. Hakala's use of multiple "dummy variables" to guard against the potential, power-reducing effects of unrelated material events lacks any basis in academic literature. Dr. Stulz's claim is incorrect.

7. Dr. Stulz claims that Dr. Hakala's procedure for identifying "material days" is arbitrary in that it "quite likely" will produce non-identical results when implemented by "separate analysts." Dr. Stulz's claim is unsubstantiated and speculative.

Basis for Opinions

Dr. Stulz's claim that Dr. Hakala's procedure is *per se* biased is incorrect.

8. Dr. Stulz states his core criticism of the Hakala event study as follows:

> As explained in detail in the [April 26, 2007, Stulz Declaration], Dr. Hakala's event study does not follow methods accepted and recognized in the finance literature. (Stulz Report, ¶ 99)
>
>> Dr. Hakala's method deviates significantly from [methods accepted and recognized in the finance literature], and biases his results towards finding more days with significant abnormal stock-price movements than he would have found had he used the conventional method. (Stulz Declaration, ¶ 25; underlining added)
>>
>> The most important aspect in which Dr. Hakala deviates from the accepted event-study method is the way in which he defines what constitutes a normal price movement and what constitutes an abnormal price movement. He subjectively selects 161 "material event" days, amounting in this case to about 38% of his sample of AOL trading days, and removes these days from his calculation of "normal" price movements for AOL …. Since any company's returns, including in this case AOL's, are likely to be more volatile on days with news than on other days, removing news days from his analysis has the effect of creating a downward bias in his estimate of the volatility of abnormal returns. This, in turn, triggers a finding of "abnormal" movement more frequently than would be triggered in the accepted methodology for event studies. (Stulz Declaration, ¶ 26; underlining added)
>
> Put simply, the most remarkable aspect of Dr. Hakala's approach is that it unfailingly inflates the statistical significance of abnormal stock returns of the subject company. The main reason that Dr. Hakala's approach invariably results in a greater number of days with larger statistical significance is because he "dummies" out a large number of days with stock-price reactions in his regression period, which results in a lower "standard error" of his regression. That is, he selects a large number of "material event days," which tend to have larger stock price reactions, and, effectively, does not include them in his regression to calculate what a normal stock-price movement would be.

> Inappropriately removing larger stock-price reactions from an analysis of normal price movements leads to an understatement of what is normal, thus causing more days in an event study to be considered normal [sic]. (Stulz Report, ¶ 100; underlining added)

9. For a proper appreciation of this criticism by Dr. Stulz of Dr. Hakala's method, it is necessary to identify clearly what is in dispute. The Hakala event study concerns, in part, whether AOL stock price movements on days with potentially material, public disclosure events concerning issues in this litigation were large enough to qualify as "statistically significant," after accounting for the likely effects on the AOL price of non-AOL-specific news having market-wide and industry-wide effects on the same days. To accomplish this, the Hakala event study uses price movement data from an "estimation period"—which should contain none of the potentially material public disclosure events in this litigation—for two related purposes: (i) to measure the degree to which AOL stock price movements typically track contemporaneous, market-wide and industry-wide price movements, and (ii) to measure the range of magnitudes of typical, AOL-specific stock price movements *not* accounted for by AOL's tracking of contemporaneous, market-wide and industry-wide price movements.

10. Item (i) provides the basis for extracting from AOL price movements on the days of potentially material AOL-specific disclosures the likely effects of non-AOL-specific news causing market-wide and industry-wide price movements on those days, thus reducing the AOL price movements to "abnormal returns" that include the effects, if any, of the corresponding, AOL-specific disclosures. Item (ii) provides the benchmark for identifying as "statistically significant" any abnormal returns that are too large to reconcile reasonably with the typical range of AOL-specific stock price movements not explained by specific, identifiable disclosures, hence

indicating that the coincident AOL-specific disclosures had non-zero effects on the price of AOL stock.

11. Dr. Hakala chose to exclude from the estimation period data used for his event study any date containing a "material event" concerning AOL.[1] Because the price movements associated with "material events" are typically larger than those on days without such disclosures,[2] this filtering of the data used to set the threshold for "statistical significance" may have the effect of lowering that threshold—as Dr. Stulz explains so emphatically. What is in dispute between Dr. Hakala and Dr. Stulz is whether such lowering of the threshold represents an appropriate and desirable increase in the sensitivity of the test for price effects of disclosures (Dr. Hakala's view) or, alternatively, imparts a "bias" that renders Dr. Hakala's results "unreliable" (Dr. Stulz's view).

12. Dr. Stulz is mistaken, in my opinion, in describing this lowering of the threshold as a "bias," and *per se* "inappropriate." Dr. Hakala views the relevant background level of price movements as representing price movements occurring on days without "material disclosures." Dr. Stulz insists, by implication, that the background level must *include* "material disclosures" unrelated to the issues in the current litigation. In Dr. Stulz's view, in other words, to be deemed

[1] Dr. Hakala outlines in ¶ 30 of his Expert Report a list of generic categories of *a priori* potentially "material events" (see footnote 14). Also in ¶ 30, he describes his procedure for identifying conforming events in the available record of public information available to investors: "[The resulting list of potentially 'material events'] was compiled through a 'blind' data selection process … without access to or reference to the actual stock price reaction on the corresponding dates."

[2] Dr. Stulz does not dispute but, rather, affirms Dr. Hakala's view that price movements associated with unrelated "material events" are typically larger than those on days without such disclosures (see the portions of the Stulz Declaration, ¶ 26, and Stulz Report, ¶ 100, quoted in my ¶ 8). From this mutually agreed premise Dr. Hakala concludes that it is proper to exclude from the baseline estimation data potentially "material events" that can be identified *a priori*, in order to improve the power of the analysis to identify "statistically significant" abnormal returns associated with disclosures related to this litigation. From the same premise Dr. Stulz draws the surprising, contrary conclusion that it is *improper* to exclude *a priori* identifiable, unrelated material events for that purpose.

"statistically significant" it is not enough for an abnormal return to outrank AOL price changes that are uncontaminated by "material events"; rather, it must outrank a mixture of such uncontaminated price changes *and* price changes that are driven by unrelated "material events."

13. It is true that any fraud-related price movement will likely stand out more prominently against Dr. Hakala's material-news-free baseline than against a baseline tainted by other, irrelevant but potentially price-moving news events. Dr. Stulz is mistaken, however, in claiming that this consequence is a "bias"; rather, this is precisely—and *properly*—the purpose of Dr. Hakala's procedure. In effect, Dr. Hakala removed some of the background "noise" from his baseline data in order to improve the sensitivity of his calculations for detecting any signal representing a price effect of fraud-related information. To claim that this more sensitive procedure is *per se* improper is incorrect: Dr. Hakala's procedure is a natural solution to the problem of contamination of baseline data by irrelevant but potentially price-moving news events. Dr. Hakala's procedure is calculated to answer, reasonably and appropriately, a question that is germane to issues in this litigation: whether AOL's "abnormal returns" coinciding with relevant, potentially material disclosures are larger than can reasonably be reconciled with AOL's typical "background" price movements *not* associated with identifiable "material events."

14. The procedure that Dr. Stulz advocates may be termed "conservative" compared to Dr. Hakala's procedure in the sense that abnormal returns that are "statistically significant" by Dr. Stulz's criterion would generally also be statistically significant by Dr. Hakala's criterion, but not necessarily vice versa. This is so because abnormal returns that appear large compared to a background that includes returns associated with "material events" will generally also appear large compared to a background from which the material-event returns have been removed. Dr.

Stulz turns this bias of his own preferred method upside down to claim that Dr. Hakala's method, lacking Dr. Stulz's "conservatism," is thereby biased and improper. This distorted characterization by Dr. Stulz is unreasonable and incorrect.

15. Moreover, Dr. Stulz supports his "bias" claim by asserting that Dr. Hakala's treatment of unrelated "material disclosures" lacks precedents in academic literature, but Dr. Stulz conspicuously fails to identify any violation by Dr. Hakala of any specific, generally accepted mathematical or statistical *principle* whose violation would logically impart the purported "bias" to the Hakala results.

Dr. Stulz's claim that Dr. Hakala's method lacks any basis in academic literature is incorrect.

16. Dr. Stulz extends his criticism of the Hakala event study as follows:

> Dr. Hakala cites a new article published in the year 2007 in the *Journal of Corporate Finance* titled "Event Studies with a Contaminated Estimation Period," by Aktas, de Bodt, and Cousin that he claims supports his approach of effectively ignoring days in the estimation of the event study. The article does nothing of the sort. The article is based on an experiment where a handful of aberrant returns are introduced into the returns of a stock over 252 days. Quite in contrast to 55% of aberrant returns, which is the number of days that Dr. Hakala dummies out in his current report, the authors consider on average less than 1% of aberrant returns. The authors focus on the statistical properties of large-scale event studies and study the properties of tests that account for event-induced variance increases. … (May 1, 2008, Stulz Report, footnote to ¶ 99; underlining added)[3]

> Dr. Hakala pursues this approach with no justification from the academic literature in a way that leads him to bias upwards the statistical significance of his results. (May 1, 2008, Stulz Report, ¶ 100; underlining added)

17. Dr. Stulz's emphatic claim that the 2007 article by Aktas, de Bodt, and Cousin (attached to this declaration as Exhibit C) lends *no* support to Dr. Hakala's method is refuted by a fair

[3] Dr. Stulz's claim that "the authors consider on average less than 1% of aberrant returns." This claim is incorrect. Section 4.2 of the article explains that the authors' simulation experiment inserted an average of 2 contaminating events, each affecting an average of 4 observations in each hypothetical estimation window of 225 observations. Thus, the true "average [proportion] of aberrant returns" was $2 \times 4/225 = 3.6\%$.

reading of the article. For example, in the following passage Aktas et al. state the problem that Dr. Hakala's method aims to solve (that is, properly taking account of "unrelated [company-specific] events" in the data used for estimation), and offer as a "natural solution" to this problem essentially the method employed by Dr. Hakala (that is, selecting estimation data "free of such contaminating events").

> The estimation period ... is most often defined as a period preceding the event ... In studies using daily data, a window going from day -250 to day -30 relative to the event date is usually (somewhat arbitrarily) chosen. This mechanical choice is, however, not free of complications. In particular, <u>unrelated events</u> may be present during the chosen estimation window, <u>which bias the estimation of the return-generating process parameters</u>. A <u>natural solution seems to be to choose, on a case-by-case basis, an estimation window free of such contaminating events</u>. This solution is, however, unreasonable for large-sample analyses. When compiling data for several hundred (or several thousand) observations (e.g., in the field of mergers and acquisitions (M&A), see Fuller et al., 2002; Mitchell and Stafford, 2000; Moeller et al., 2003), using such a "brute force" approach quickly becomes intractable. (Aktas et al., p. 130; underlining added)

Note that this passage does not reject the "natural solution" as flawed due to bias but, rather, as impracticable in the case of an event study involving hundreds or thousands of subject firms (in contrast with Dr. Hakala's event study, which involves only AOL). Indeed, Aktas et al. warn here that bias may result from *failing* to control for "unrelated events" in the estimation data, as Dr. Stulz failed to do in his own event study described in his Declaration, and as Dr. Stulz implicitly recommends in his Expert Report. This is precisely the form of bias that Dr. Hakala's method aims to *avoid*.

18. Dr. Stulz attempts to distinguish Dr. Hakala's procedure from the examples discussed by Aktas et al. on the basis that the *proportion* of potentially contaminated observations identified by Dr. Hakala is much greater than the proportions used by Aktas et al. in their simulation experiments. However, Dr. Stulz conspicuously fails to identify any specific mathematical or

statistical principle under which this superficial distinction would imply any essential logical gap between Dr. Hakala's procedure and the considerations set forth by Aktas et al. Aktas et al. state no explicit limitation of the scope of their comments to the case of only a small proportion of contaminating events, and no such limitation is logically implied by their analysis of the issue.

19. Moreover, the article by Aktas et al. is not the only precedent in academic literature for Dr. Hakala's procedure. For example, in an article published in 1988 in the Journal of Business Finance & Accounting (and attached to this declaration as Exhibit D), Dr. Joel E. Thompson—then a faculty colleague of Dr. Stulz's at The Ohio State University—reported a study of three potential refinements of event study methods.[4] He describes the third of these refinements as follows.

> In addition, the importance of extraneous individual firm events occurring during the estimation period is investigated. These extraneous events may increase the variance of a firm's returns resulting in a larger estimated variance and thereby decrease the power of the test. Hence, returns associated by time with such events during the estimation period are deleted to determine whether this procedure results in a more powerful test. (Thompson (1988), p. 78)
>
> … [T]he importance of extraneous individual firm events is examined by using firm events cited in the *Wall Street Journal Index* (*WSJI*). Individual firm returns occurring the trading day before, the trading day of, or the trading day after a firm event cited in the *WSJI* were deleted from that firm's estimation period. … Examination of [the results of this analysis] reveals that [the statistical] power [i.e., the sensitivity of the method] can be improved by using such a partial estimation period [i.e., by deleting returns associated with disclosures cited in the *WSJI*] …. (Thompson (1988), pp. 81 and 83; underlining added)

[4] Joel E. Thompson, "More Methods that Make Little Difference in Event Studies," Journal of Business Finance & Accounting, 15(1) Spring 1988.

The method described here is obviously a close parallel to the Hakala procedure that Dr. Stulz claims lacks precedents in academic literature. Moreover, Dr. Thompson's conclusion supports Dr. Hakala's stated reason for using this procedure in the present case.[5]

Dr. Stulz offers no specific basis for his claim that Dr. Hakala's procedure for identifying "material days" cannot be replicated "exactly," and no evidence that differences among reasonable alternative implementations would have any material effect on Dr. Hakala's key results.

20. Dr. Stulz concludes his core criticism of the Hakala event study as follows:

> As another indication of the arbitrariness of this methodology, consider that Dr. Hakala's choice of material days cannot be replicated exactly by an independent scientist, that is, two separate analysts are quite likely to come up with different results when implementing his methodology. Further, his choice of material events is at times extremely surprising. … (Stulz Report, ¶ 101)

However, Dr. Stulz provides no examples of differences among reasonable implementations of Dr. Hakala's method by "independent scientists" and demonstrates no material effects of any such differences on Dr. Hakala's findings. Lacking such foundation, this Stulz opinion is mere speculation.

21. The opinions I have reached in this case are based on information, data, and analyses of the types typically and reasonably relied upon by experts in statistics and applied mathematics.

---

[5] Dr. Thompson's conclusion concerning all three of the methodological refinements considered in his study is that they "make little difference": "it was found in this study that industry indexes, return form, and individual firm extraneous events have little impact on event study results. … [I]n general, the factors examined in this study do not appear to have an important role in event studies" (p. 84). Accordingly, he does not recommend that any of them—including, in particular, Dr. Hakala's procedure—be adopted for general use in event studies. Note, however, that this is only because these methods "make little difference" in Dr. Thompson's simulation study, not because of any logical defect or "bias." In particular, Dr. Thompson concludes that a refinement like Dr. Hakala's procedure can improve the statistical sensitivity of an event study but that "the improvement is small" (p. 83). (Note that the focus of Dr. Stulz's criticism is not the *size* of any improvement in sensitivity, but the basic logical validity of Dr. Hakala's procedure.)

Executed this 16th day of July 2008 in Novato, California. I declare under penalty of perjury that the foregoing is true and correct.

M. L. Marais

M. Laurentius Marais

Exhibit A

Curriculum Vitae of M. Laurentius Marais

July 2008

M. LAURENTIUS MARAIS

505 San Marin Drive
Novato, California 94945

Tel: (415) 898-2255
Fax: (415) 898-2260

Web: www.wecker.com

EDUCATION:

Ph.D.	Stanford University (Business Administration, Mathematics), 1985
M.S.	Stanford University (Statistics), 1983
M.S.	Stanford University (Mathematics), 1976
B.Sc.	Stellenbosch University (Mathematics, Applied Mathematics, Computer Science), 1973

EMPLOYMENT:

1993 to date	Vice President, William E. Wecker Associates, Inc.
1994-1998	Stanford University, Consulting Professor, School of Law
1992 to date	Senior Consultant, now Principal Consultant, William E. Wecker Associates, Inc.
1982-1991	University of Chicago, Instructor, later Assistant and Associate Professor, Graduate School of Business.

ACTIVITIES:

Editorial Board, Journal of Accounting Research, 1987-1992

Refereed for:
- The Accounting Review
- Contemporary Accounting Research
- Journal of Accounting and Economics
- Journal of Accounting Research
- Journal of Business and Economic Statistics
- Journal of Financial Research
- Journal of Money, Credit and Banking

Member of:
- American Accounting Association
- American Economic Association
- American Statistical Association
- Royal Statistical Society
- Mathematical Association of America
- Society for Industrial and Applied Mathematics

PUBLICATIONS and WORKING PAPERS:

"The experimental design of classification models: an application of recursive partitioning and bootstrapping to commercial bank loan classifications," (with James M. Patell and Mark A. Wolfson), Journal of Accounting Research, 1984.

"An application of the bootstrap method to the distribution of squared, standardized market model prediction errors," Journal of Accounting Research, 1984.

"An analysis of a multivariate regression model in the context of a regulatory event study by computer intensive resampling," Working Paper, Institute of Professional Accounting, University of Chicago, July 1986.

"A note on the algebraic and statistical properties of the multivariate market model," Working Paper, Institute of Professional Accounting, University of Chicago, September 1986.

"On drawing inferences about market reactions to the regulation of accounting for oil and gas exploration: An application of computer intensive resampling methods," Working Paper, Institute of Professional Accounting, University of Chicago, September 1986.

"On detecting abnormal returns to a portfolio of nonsynchronously traded securities," Working Paper, Institute of Professional Accounting, University of Chicago, October 1986.

"Reduced demands on recovery room resources with Diprivan compared to thiopental-isoflurane," (with Michael W. Maher et al.), Anesthesiology Review, January/February 1989.

"Wealth effects of going private for senior securities," (with Katherine Schipper and Abbie J. Smith), Journal of Financial Economics, 1989.

"Consequences of going-private buyouts for public debt and preferred stock: 1974 to 1985," (with Katherine Schipper and Abbie J. Smith), in Proceedings of the 25th Annual Conference on Bank Structure and Competition: Banking System Risk - Charting a New Course, Federal Reserve Bank of Chicago, 1989.

"Discussion of 'Post-earnings-announcement drift: Delayed price response or risk premium?'," Journal of Accounting Research, 1989.

"Using relative productivity assessments for allocating housestaff to departments," (with Michael W. Maher, Michael F. Roizen, et al.), Medical Care, 1990.

"An adaptable computer model of the economic effects of alternative anesthetic regimens in outpatient surgery," (abstract; with Michael W. Maher et al.), Anesthesiology (Supplement), September 1990.

"On the finite sample performance of estimated generalized least squares in seemingly unrelated regressions: nonnormal disturbances and alternative standard error estimators," Working Paper, Institute of Professional Accounting, University of Chicago, January 1991.

"Exploiting tax attributes of spinoffs to structure takeovers and takeover-related defenses," (with Katherine Schipper), Working Paper, Institute of Professional Accounting, University of Chicago, August 1991.

"Technological innovation and firm decision-making: accounting, finance and strategy," (with Paul J. H. Schoemaker), Working Paper, Institute of Professional Accounting, University of Chicago, September 1991.

"Process-oriented activity-based costing," (with Michael W. Maher), Working Paper, Institute of Professional Accounting, University of Chicago, June 1992.

"A field study on the limitations of activity-based costing when resources are provided on a joint and indivisible basis" (with Michael W. Maher), Journal of Accounting Research, 1998.

"Correcting for omitted-variables and measurement-error bias in regression with an application to the effect of lead on IQ" (with William E. Wecker), Journal of the American Statistical Association, June 1998.

"Event study methods: detecting and measuring the security price effects of disclosures and interventions" (with Katherine Schipper), in Litigation Services Handbook: The Role of the Financial Expert, 2005 Cumulative Supplement, 3rd Ed., John Wiley & Sons.

"Estimating Cost Behavior" (with Michael W. Maher), in Handbook of Cost Management, 2005, 2nd Ed., John Wiley & Sons.

"Audit Committee Financial Literacy: A Work in Progress" (with Douglas J. Coates and Roman L. Weil), March 2005. CRSP Working Paper No. 605. Available at SSRN: http://ssrn.com/abstract=680281

"Statistical Estimation of Incremental Cost from Accounting Data" (with Michael W. Maher, William E. Wecker, and Roman L. Weil), in Litigation Services Handbook: The Role of the Financial Expert, 2006, 4th Ed., John Wiley & Sons.

Exhibit B

Documents Received and Reviewed

Documents Received and Reviewed

1) "Expert Report of Scott D. Hakala, Ph.D, CFA," dated March 4, 2008.

2) "Expert Report of René M. Stulz," dated May 1, 2008.

3) "Declaration of René M. Stulz," dated April 26, 2007.

Exhibit C

Aktas, de Bodt, and Cousin (2007)

ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect

Journal of Corporate Finance 13 (2007) 129 145

Journal of CORPORATE FINANCE

www.elsevier.com/locate/jcorpfin

Event studies with a contaminated estimation period

Nihat Aktas [a,*], Eric de Bodt [a,b], Jean-Gabriel Cousin [b]

[a] *IAG Louvain School of Management, Université catholique de Louvain, Belgium*
[b] *ESA, Université de Lille 2, France*

Received 14 September 2005; received in revised form 13 September 2006; accepted 27 September 2006
Available online 13 November 2006

Abstract

Event studies are an important tool for empirical research in Finance. Since the seminal contribution of Fama et al. [Fama, E., Fisher, L., Jensen, M., Roll, R., 1969. The adjustment of stock prices to new information. International Economic Review 10, 1 21], there have been many enhancements to the classical test methodology. Somewhat surprisingly, the estimation period has attracted less interest. It is usually routinely determined as a fixed window prior to the event announcement day. In this study, we propose a test that reduces the impact of potentially unrelated events during the estimation period. Our proposition is based on a two state version of the classical market model as a return generating process. We present standard specification and power analyses. The results highlight the importance of explicitly controlling for unrelated events occurring during the estimation window, especially in the presence of event induced increase in return volatility.

JEL classification: G14; G34
Keywords: Event study; Unrelated events; Markov switching regression model

1. Introduction

Since the seminal contribution of Fama, Fisher, Jensen and Roll (1969) (hereinafter referred to as FFJR), event studies have become a standard empirical methodology in research in Finance. Applications are so numerous that it is impractical to try to list them exhaustively. Many suggestions have been put forward to improve the basic empirical methodology. Brown and Warner (1980, 1985) analyzed the specification and power of several modifications of the FFJR

* Corresponding author. Place des Doyens 1, B-1348 Louvain-la-Neuve, Belgium. Tel.: +32 10 478439; fax: +32 10 478324.
E-mail address: aktas@fin.ucl.ac.be (N. Aktas).

0929-1199/$ - see front matter
doi:10.1016/j.jcorpfin.2006.09.001

approach. Ball and Torous (1988) explicitly took into account the uncertainty about event dates. Corrado (1989) introduced a non-parametric test of significance. Boehmer et al. (1991) proposed an adaptation of the standard methodology to tackle an event-induced increase in return volatility. Since this contribution, most methodological papers have explicitly controlled for this important phenomenon. Salinger (1992) suggested an adjustment of the abnormal returns standard errors robust to event clustering. Savickas (2003) recommended the use of a GARCH specification to control for the effect of time-varying conditional volatility. Aktas et al. (2004) advocated the use of a bootstrap method as an alternative to Salinger's (1992) proposition. Recently, Harrington and Shrider (in press) have argued that all events induce variance, and therefore tests robust to cross-sectional variation should always be used.

The estimation period has attracted less attention. It is most often defined as a period preceding the event, which is sufficiently long to enable the parameters of the chosen return-generating process to be properly estimated. In studies using daily data, a window going from day −250 to day −30 relative to the event date is usually (somewhat arbitrarily) chosen. This mechanical choice is, however, not free of complications. In particular, unrelated events may be present during the chosen estimation window, which bias the estimation of the return-generating process parameters. A natural solution seems to be to choose, on a case-by-case basis, an estimation window free of such contaminating events. This solution is, however, unreasonable for large-sample analyses. When compiling data for several hundred (or several thousand) observations (e.g., in the field of mergers and acquisitions (M&A), see Fuller et al., 2002; Mitchell and Stafford, 2000; Moeller et al., 2003), using such a "brute force" approach quickly becomes intractable.

It is worth emphasizing that, in many research areas, the presence of contaminating events during the estimation window is not just a presumption. Let us take the case of M&As again. Imagine that a specific bidder has, during the months preceding the transaction being studied, undertaken other operations, as frequently it appears to be the case (see, e.g., Asquith et al., 1983; Schipper and Thompson, 1983; Malatesta and Thompson, 1985; Fuller et al., 2002; Aktas et al., 2006). For example, out of the 4135 deals comprising the M&A sample used by Fuller et al. (2002),[1] 2721 (66%) would have been contaminated if the classical definition of the estimation window had been used.[2] The existence of such firm-specific events in the estimation window will most likely affect the estimation of the return-generating process and, in particular, the estimated variance of the parameters.

The approach we introduce in this paper to solve this contaminating event problem is essentially based on a combination of the well-established market model (Sharpe, 1963) and the more recent Markov switching regression models, largely introduced and developed by Hamilton (1989, 1994) and significantly extended by Krolzig (1997). Using a two-state market model, the estimated parameters of the model are less subject to the influence of contaminating events. This can be understood as a statistical filtering of the data. Another way to interpret our proposition is to see it as a better-specified return-generating model, which takes into account the probability of the occurrence of firm-specific events. From this perspective, our approach is in line with Roll's (1987) results. According to Roll, the true return-generating process seems to be better described by a mixture of two distributions: one corresponding to a state of information arrival, and the other to the normal return behavior.

[1] We thank the authors for providing us with access to their data set.

[2] This is the reason why the authors use the BETA-1 return-generating process, which is not affected by unrelated events during the estimation window, since there are no parameters to be estimated. However, according to our analyses, this approach is clearly less powerful than the other alternatives (see Section 5).

The analysis that we develop in this paper is now classical in the field of event study methodology (Brown and Warner, 1980, 1985). Using daily CRSP data, we carried out specification and power analyses while simulating a contaminated estimation period. We compare our approach to a classical set of alternatives (such as Corrado (1989), Boehmer et al. (1991), Savickas (2003)). The results show that (i) our approach is robust to the estimation window contamination and that (ii), in the context of an event-induced increase in return volatility, it dominates competing methods. Given the results of Harrington and Shrider (in press), following which event-induced increase in return volatility must be taken into account, we recommend the use of our approach.

The paper is organized as follows: Section 2 presents a simple model to show that ordinary least squares (OLS) methods overestimate the standard error of an individual firm's abnormal return when the true process is state dependent. Section 3 is devoted to a short review of the classical event study approaches and to the presentation of our test. Section 4 describes our experimental design. In Section 5, we present simulation results comparing the specification and the power of the test statistics being considered. Section 6 contains our summary and conclusions.

2. State dependent return-generating process and OLS inferences

In this section we show that OLS estimators overestimate the standard error of an individual firm's abnormal returns when the true return-generating process has two-states. We use the following notation:

- $\mathbf{X}_j$ is the matrix of explanatory variables for firm j;
- $\boldsymbol{R}_j$ and $\boldsymbol{R}_m$ are vectors of returns for firm j and for a market portfolio proxy;
- D_j is a dummy variable equal to 1 at the event date for firm j, and 0 otherwise;
- $\boldsymbol{b}_j$ is the vector of coefficient estimates for firm j.

When estimating firm j's abnormal returns using the market model (MM) as the return-generating process (Sharpe, 1963), $\mathbf{X}_j$ has three columns (the first column is set to 1 for the constant, the second to $\boldsymbol{R}_m$ and the third to D_j). The vector of coefficients $\boldsymbol{b}_j$ is composed of the intercept of MM, the β coefficient for firm j and a third coefficient (denoted by γ) capturing the firm j's abnormal returns at the announcement date:

$$\boldsymbol{R}_j = \mathbf{X}_j\boldsymbol{b}_j + \varepsilon_j = [1\boldsymbol{R}_m D_j]\begin{bmatrix}\alpha\\ \beta\\ \gamma\end{bmatrix} + \varepsilon_j. \tag{1}$$

Assuming homoskedascity, the covariance matrix of the OLS estimator is well known:

$$\mathrm{COV}_j^{\mathrm{OLS}}(\boldsymbol{b}_j|\mathbf{X}_j) = \sigma_j^2(\mathbf{X}_j{}'\,\mathbf{X}_j)^{-1}. \tag{2}$$

Assume now that the residuals are state dependent. S_t denotes the state variable. We consider the case of a two-state regime model. More precisely, we have a low variance regime (S_t=1) and a high variance regime (S_t=2):[3]

$$\begin{aligned}\boldsymbol{R}_j &= \mathbf{X}_j'\,\boldsymbol{b}_j + \varepsilon_{j,1} \quad \text{if } S_t = 1\\ \boldsymbol{R}_j &= \mathbf{X}_j'\,\boldsymbol{b}_j + \varepsilon_{j,2} \quad \text{if } S_t = 2.\end{aligned} \tag{3}$$

[3] Note that the market-model parameters $\boldsymbol{b}_j$ are identical in the two states. However, this setup could be easily extended to take regime dependent parameters into account.

The variance of the residuals for each state is set to:

$$\begin{aligned} E[\varepsilon_{j,1}\varepsilon_{j,1}'|X] &= \sigma_{j,1}^2 I \quad \text{if } S_t = 1 \\ E[\varepsilon_{j,2}\varepsilon_{j,2}'|X] &= \sigma_{j,2}^2 I \quad \text{if } S_t = 2' \end{aligned} \tag{4}$$

where $\sigma_{j,2}^2 > \sigma_{j,1}^2$. The high variance regime allows us to explicitly incorporate the presence of unrelated events into the statistical model.

The transition between the two regimes is governed by a Markov chain of order 1, for which the transition matrix is given by:

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1-p_{22} \\ 1-p_{11} & p_{22} \end{bmatrix}, \tag{5}$$

where $p_{m,n} = p(S_t = m | S_{t-1} = n)$ corresponds to the probability of changing from state n to state m. The unconditional probability of the regime is given by (Hamilton, 1994, p. 683):

$$p(S_t = 1) = \frac{1-p_{22}}{2-p_{11}-p_{22}};\, p(S_t = 2) = \frac{1-p_{11}}{2-p_{11}-p_{22}}. \tag{6}$$

Assuming strict exogeneity and that the transition probabilities are deterministic, the covariance matrix of the OLS estimator is a weighted average of the covariance matrix in the two-states, the unconditional probability being the weights:

$$\mathrm{COV}^{\mathrm{OLS}}(\boldsymbol{b}_j|\mathbf{X}_j) = p(S_t = 1)\sigma_{j,1}^2(\mathbf{X}_j'\mathbf{X}_j)^{-1} + p(S_t = 2)\sigma_{j,2}^2(\mathbf{X}_j'\mathbf{X}_j)^{-1}. \tag{7}$$

Considering the null hypothesis of no event, we are interested in the low variance regime covariance matrix, which is $\sigma_{j,1}^2(\mathbf{X}_j'\mathbf{X}_j)^{-1}$, and since $\sigma_{j,2}^2 > \sigma_{j,1}^2$, this leads to the inequality

$$\sigma_{j,1}^2(\mathbf{X}_j'\mathbf{X}_j)^{-1} \leq p(S_t = 1)\sigma_{j,1}^2(\mathbf{X}_j'\mathbf{X}_j)^{-1} + p(S_t = 2)\sigma_{j,2}^2(\mathbf{X}_j'\mathbf{X}_j)^{-1}. \tag{8}$$

Eq. (8) shows that the standard error of the OLS estimates of the abnormal return (γ_j) is overestimated when the true generating process is a two-state process. This result offers a clear econometric foundation for the loss of power of classical event study methodology when the estimation window is contaminated by unrelated events.

3. Event study methodology

The seminal contribution of FFJR has been the starting point for an impressive diffusion of event study methodology in finance, accounting and economics. Its component steps are well known. In this section, we focus on the choice of the return-generating process and the construction of the statistical test of significance, two key points of concern in this area. The set of approaches on which we focus has been chosen either because they are used in classical empirical studies (e.g., the standardized cross-sectional test (Boehmer et al., 1991) and the RANK test (Corrado, 1989)) or because they contain features making them potentially well-suited to controlling for the presence of unrelated events during the estimation period (e.g., the BETA-1 approach and the GARCH approach developed by Savickas (2003)). Many other propositions

have been made in the literature (e.g., Ball and Torous, 1988; Nimalendran, 1994) but they do not relate directly to our work.

3.1. Return generating processes

Abnormal returns (AR) correspond to the forecast errors of a specific normal return-generating model (in Section 2, to simplify the exposition, we expressed them as the coefficient values of a dummy variable). Using MM as return-generating process[4] and employing the notation described in Section 2, we obtain

$$\boldsymbol{R}_j = \alpha_j + \beta_j \boldsymbol{R}_m + \varepsilon_j. \tag{9}$$

The residuals, ε_j, provide the estimates of abnormal returns. Classically, the residuals are supposed to be identically and independently (normally) distributed (IID). Numerous contributions have dealt with violations of this hypothesis. For example, Ruback (1982) suggests a way of coping with the existence of first-order auto-correlation in asset returns.

3.2. Statistical tests of significance

To introduce the different tests to be analyzed, we expand the notation used in Section 2 following Boehmer et al. (1991):

- N: number of firms in the sample;
- AR_{jE}: abnormal return of firm j on the event date;
- AR_{jt}: abnormal return of firm j on date t;
- T: number of days within the estimation period;
- TE: number of days within the event period;
- $\bar{\boldsymbol{R}}_m$: average return on the market portfolio during the estimation period;
- $\hat{S}_j$ standard deviation of firm j's AR during the estimation period;
- SR_{jE}: standardized AR of firm j on the event date, calculated as:

$$\mathrm{SR}_{jE} = \mathrm{AR}_{jE} \Bigg/ \left[\hat{S}_j \sqrt{1 + \frac{1}{T} + \frac{(R_{m,E} - \bar{R}_m)^2}{\sum_{t=1}^{T} (R_{m,t} - \bar{R}_m)^2}} \right]. \tag{10}$$

3.2.1. The BMP test

For each of the significance tests, we consider the null hypothesis of no cross-sectional average (cumulative) abnormal returns around the event date. The BMP (Boehmer, Musumeci and Poulsen) test is similar in spirit to Patell's (1976) test. However, Boehmer et al. (1991) use the estimated cross-sectional variance of the standardized abnormal returns instead of the theoretical

[4] Brown and Warner (1980, 1985) show that the results of short-term event studies are not sensitive to the choice of a specific return-generating process. Aktas et al. (2004) reported similar results in a European context.

value. This adaptation captures the event-induced increase in return volatility. The BMP test takes the following form:

$$Z_{\mathrm{BMP}} = \frac{\frac{1}{N}\sum_{j=1}^{N} \mathrm{SR}_{jE}}{\sqrt{\frac{1}{N(N\ \ 1)}\sum_{j=1}^{N}\left(\mathrm{SR}_{jE} - \sum_{i=1}^{N}\frac{\mathrm{SR}_{iE}}{N}\right)^2}}. \tag{11}$$

3.2.2. The BETA 1 test

The BETA-1 test may be viewed as a drastic simplification of the BMP test. The assumed return-generating process is the market-adjusted model, which amounts to imposing $\beta=1$ and $\alpha=0$ in Eq. (9). The test is based on cross-sectional estimates of the standard deviation of the event-day abnormal returns (ARE). We therefore obtain:

$$Z_{\mathrm{BETA}\ \ 1} = \frac{\frac{1}{N}\sum_{j=1}^{N} \mathrm{AR}_{jE}}{\sqrt{\frac{1}{N(N\ \ 1)}\sum_{j=1}^{N}\left(\mathrm{AR}_{jE} - \sum_{j=1}^{N}\frac{\mathrm{AR}_{jE}}{N}\right)^2}}. \tag{12}$$

The BETA-1 test constitutes, in our framework, an interesting alternative. $Z_{\mathrm{BETA\text{-}1}}$ does not use data from the estimation window and is therefore free of the potential biases generated by contaminating events.

3.2.3. The Corrado (1989) RANK test

Corrado (1989) introduced a test based on the ranks of abnormal returns. The RANK test merges the estimation and event windows in a single time series. Abnormal returns are sorted and a rank is assigned to each day. If K_{jt} is the rank assigned to firm j's abnormal return on day t, then the RANK test is given by

$$T_{\mathrm{CORRADO}} = \frac{\frac{1}{N}\sum_{j=1}^{N}(K_{jE} - K)}{S(K)}, \tag{13}$$

where K is the average rank and $S(K)$ is the standard error, calculated as

$$S(K) = \sqrt{\frac{1}{T+TE}\sum_{t=1}^{T+TE}\left(\frac{1}{N}\sum_{j=1}^{N}(K_{jt} - K)\right)^2}. \tag{14}$$

The use of ranks neutralizes the impact of the shape of the AR distribution (e.g., its skewness and kurtosis and the presence of outliers). It should therefore represent an attractive alternative

way of neutralizing contaminating events within the estimation window. Corrado (1989), Corrado and Zivney (1992), and Campbell and Wasley (1993) provide an in-depth analyses of this approach.

3.2.4. The GARCH test

The conditional time-varying behavior of the variance of returns has been widely recognized in finance since it was first pointed out by Engle (1982). Building on the Bollerslev (1986) generalized autoregressive conditional heteroskedastic (GARCH) approach, Savickas (2003) suggested the use of the return-generating process

$$\begin{aligned} \boldsymbol{R}_{j,t} &= \alpha_j + \beta_j \boldsymbol{R}_{m,t} + \gamma_j D_{j,t} + \eta_{j,t} \\ \eta_{j,t} &\sim N(0, h_{j,t}), \\ h_{j,t} &= aj + b_j h_{j,t\ 1} + c_j \eta^2_{j,t\ 1} + d_j D_{j,t} \end{aligned} \tag{15}$$

where $h_{j,t}$ is the conditional time-varying variance and a_j, b_j, c_j and d_j are the coefficients of the GARCH(1,1) specification. $D_{j,t}$ is a dummy variable equal to 1 at the event date for firm j, and 0 otherwise. As in Section 2, the γ_j coefficient captures the abnormal return at the announcement date.

The conditional variance $h_{j,t}$ provides a natural estimator of the AR variance. Savickas (2003) used it to standardize the AR before proceeding with the BMP test. In this setting, Eq. (10) is replaced by

$$\mathrm{SR}_{jE} = \hat{\gamma} j \Big/ \sqrt{\hat{\mathrm{h}}_{j,E}}. \tag{16}$$

Savickas (2003) shows that the GARCH test allows him to control for the time-varying variance of AR and the event-induced increase in return volatility. It is worth mentioning that the simulation results he provides rely on abnormal returns generated on five consecutive days. This very specific simulation procedure must be kept in mind when interpreting the specification and power results that we present in Section 5. As the main effect of contaminating events is an increase in variance during the estimation window, the GARCH test also constitutes an attractive alternative to the BMP test. The GARCH(1,1) specification might indeed provide a way of controlling for the impact of unrelated events on the estimate of the variance in AR (probably at the cost of an increase in persistence).

3.2.5. The two state market model test (TSMM)

The idea underpinning the TSMM test is simple: the presence of unrelated events within the estimation window has an impact on the (*ex post*) estimation of the AR variance. Classical tests, such as the BMP, will overestimate the variance of the residuals during the estimation period in such conditions, leading to a downward bias in the significance test (i.e. less likelihood of rejecting the null hypothesis) during the event window.

To deal with this bias, the TSMM test relies on the Markov switching regression framework developed by Hamilton (1989, 1994). As discussed in Section 2, we assume that the return-generating process can be adequately modeled by a two-state process,[5] in which one regime has

[5] This hypothesis is supported by unpublished results. Using the approach developed by Krolzig (1997), we found that a two-state model was an adequate representation of the return-generating process in most cases. Three-regime decomposition appeared to be justified only in the presence of strong outliers.

normal variance and the other high variance. Note that the MM parameters are assumed to be the same in the two regimes.[6] Eq. (3) then becomes

$$\begin{aligned} \boldsymbol{R}_{j,t} &= \alpha_j + \beta_j \boldsymbol{R}_{m,t} + \gamma_j D_{j,t} + \varepsilon_{j,S,t} \\ \varepsilon_{j,S,t} &\sim N(0, \sigma^2_{j,s}), \end{aligned} \tag{17}$$

where S is a state variable taking value 1 in the low variance state and value 2 in the high variance state ($\sigma^2_{j,2} > \sigma^2_{j,1}$). The proposed model is a direct and parsimonious extension of the classical MM. As for the previous GARCH model, the γ_j coefficient corresponds to the estimated event-day abnormal return. We use the estimated standard error of γ_j to standardize the AR. Our standardized abnormal return is therefore

$$\mathrm{SR}_{jE} = \hat{\gamma}_j / \mathrm{SE}(\gamma_j), \tag{18}$$

where $\mathrm{SE}(\gamma_j)$ corresponds to the standard error of the γ_j coefficient. The test can then be constructed using the same approach as for ZBMP:

$$Z_{\mathrm{TSMM}} = \frac{\frac{1}{N}\sum_{j=1}^{N} \mathrm{SR}_{jE}}{\sqrt{\frac{1}{N(N\ 1)}\sum_{j=1}^{N}\left(\mathrm{SR}_{jE} - \sum_{i=1}^{N}\frac{\mathrm{SR}_{iE}}{N}\right)^2}} \tag{19}$$

The estimation of Eq. (17) is based on a maximum-likelihood approach.[7] The estimated probability of being in a specific state on a specific date is one of the interesting byproducts of this approach. In some specific cases, it allows the reasons for an increase in variance to be explored.[8]

4. Experimental design

Our investigation of the specification and power of the TSMM test follows the procedure introduced by Brown and Warner (1980, 1985) and used repeatedly since then (see, e.g., Corrado, 1989; Boehmer et al., 1991; Corrado and Zivney, 1992; Cowan, 1992; Cowan and Sergeant, 1996; Savickas, 2003).

4.1. Data and sample selection

Our universe of firms is composed of companies included in the Center for Research in Security Prices (CRSP) daily returns file from January 1, 1973 through December 31, 2004.[9] As

[6] We have also implemented a more general version of the TSMM approach, where the MM parameters are regime dependent. Since the results are similar, we have not reported them in this paper. The assumption that the MM parameters are the same for both regimes does not rely on a specific economic/financial foundation. It is simply a product of the principle of parsimony, which allows for much more efficient numerical estimation. We thank an anonymous referee for having stressed this point.

[7] All the estimations presented in this paper have been performed with the Ox econometric software, using the Krolzig MSVAR package. We thank Professor J. Hamilton for advising us on the use of this package.

[8] In a recent paper, Cousin and de Launois (2006) produce empirical evidence supporting that the high variance state is generated by intensive information arrival.

[9] We chose to start the sampling process in 1973 in order to draw the samples from an homogenous population, Nasdaq stocks being only included in the CRSP files from 1973 onwards.

our market portfolio we used the CRSP value weighted index. All firms and event dates were randomly chosen with replacement such that each firm/date combination had an equal chance of being chosen at each selection. For each replication, we constructed 1000 samples of 50 firms. The estimation window length was 225 days and the event date was situated at day 250. Like Savickas (2003), our sampling process excluded securities with missing returns during the 250-day interval. Moreover, to be included in the samples, securities need to have at least 100 non-zero returns over the estimation window, and no zero return due to a 'reported price' on the event-day.

4.2. Contaminating the estimation window

The aim of our simulation was to study the specification and power of the TSMM test, as compared to other tests, when the data in the estimation window were deliberately contaminated. To simulate significant events, we injected AR into the estimation window (on a contaminated day) which was twice the standard deviation of the actual stock. To generate stochastic shocks, we followed the method proposed by Brown and Warner (1985) by adding another two demeaned returns randomly drawn from the estimation window. Moreover, we generated both positive and negative AR to represent the unknown nature of the events likely to affect the estimation window. The sign (±) of the simulated AR ($2^*\sigma_{\boldsymbol{R}}$) was determined by random sampling from a Bernoulli distribution. The transformed return for security j on a contaminated day, denoted $\boldsymbol{R}'_{j,t}$, is therefore computed as

$$\boldsymbol{R}'_{j,t} = \boldsymbol{R}_{j,t} \pm 2^* \sigma_{\boldsymbol{R}_j} + (\boldsymbol{R}_{j,X} - \boldsymbol{R}_j) + (\boldsymbol{R}_{j,Y} - \boldsymbol{R}_j), \tag{20}$$

where $\boldsymbol{R}_{j,t}$ is the actual return, $\boldsymbol{R}_{j,X}$ and $\boldsymbol{R}_{j,Y}$ are returns randomly selected from the estimation period, and $\boldsymbol{R}_j$ is the average return in the estimation period.

The number and nature of the events during the estimation window was determined in two steps. First, a random sample was drawn from a Poisson distribution with a mean of two. This value, denoted by λ^*, represents the number of events during the estimation window. Events were then randomly assigned to specific days in the estimation window (by random sampling from a uniform distribution). Finally, the length (in number of days) of each event was again randomly sampled from a Poisson distribution, this time with a mean of four. This approach allowed random events to be generated in the estimation window.

Fig. 1 presents a typical result. The solid line is the time series of the initial returns, and the dotted line is the time series of the returns obtained after the generation of random, contaminating events. Three events were generated, at dates $T=2$, 51 and 119. The standard error of the initial estimation period was 3.73%, and after the generation of events it became 4.01%. Fig. 1 shows that the simulated shocks are not particularly uncommon events. They merely represent disruptions in the normal return-generating process, of the size and amplitude we might expect from various corporate event announcements (e.g., M&A operations, share buy-backs, earning announcements).

4.3. Simulating abnormal performance

We generated abnormal returns at the event date in the same way as Brown and Warner (1980 1985) by adding a constant to each stock return observed on day 0 (event date). The abnormal

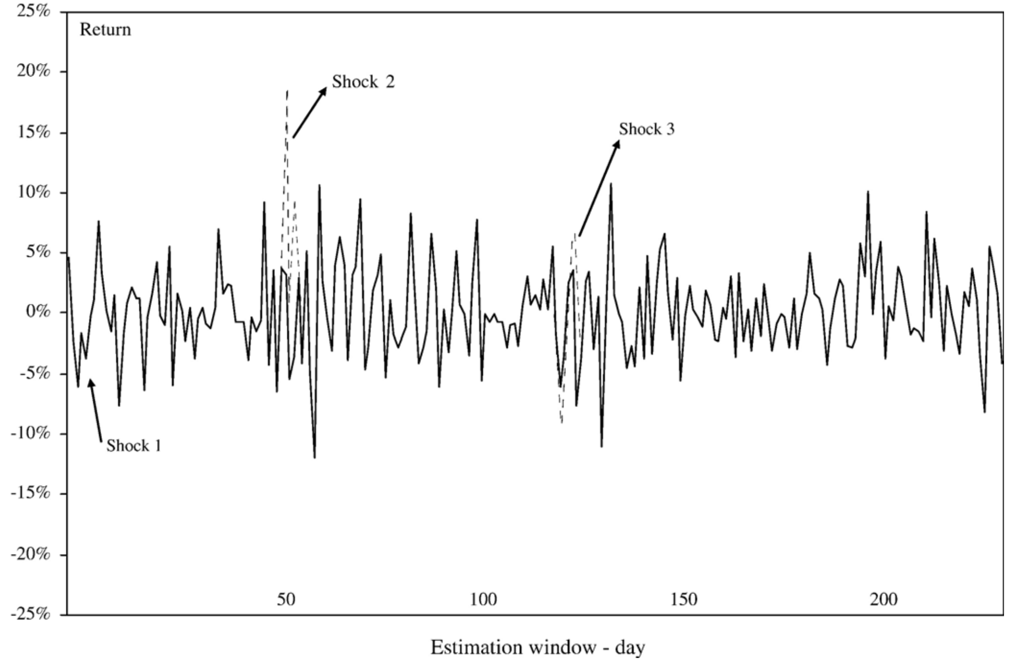

Fig. 1. This figure provides an example of estimation window contamination using the procedure described in Section 4.2. The solid line is the time series of initial returns and the dotted line is the time series obtained after event generation. Three events are generated, at dates T 2, 51 and 119. The standard error of the estimation window is 3.73% initially and 4.01% after event generation.

performance simulated is 0% for the specification analysis and +1% for the power analysis.[10] To produce stochastic abnormal returns (the event-induced variance phenomenon) we again followed Brown and Warner's approach. Each security's day 0 return, $R_{j,0}$, was transformed to triple its variance by adding two demeaned returns randomly drawn from the estimation window. The event-day transformed return was obtained using the procedure described in Section 4.2 above.

5. Empirical results

Our results are presented in Tables 1–4, which show the rejection rates for different cross-sectional test statistics under different conditions. We compare specifications (Tables 1 and 3) and powers (Tables 2 and 4) without (Tables 1 and 2) and with (Tables 3 and 4) event-induced increase in volatility.[11] In each case, we present the results without (Panel A) and with (Panel B) contaminating events during the estimation window. Results are presented for the BMP, RANK, GARCH, BETA-1 and TSMM tests.

[10] Unreported results show that, beyond a simulated abnormal performance of 2%, all the methods we compared are very powerful.

[11] According to Harrington and Shrider (in press) all events induce variance. We have retained in the paper all the analyses of the "no event-induced variance" cases (Tables 1, 2) to allow comparison with previous papers in the field.

Table 1
Rejection rates of test statistics: no event-induced returns no event-induced variance

	Significance level		
	1%	5%	10%
Panel A. Without contaminating events			
BMP	1.60%	5.20%	9.80%
RANK	0.90%	4.50%	8.40%
GARCH	0.70%	5.90%	11.40%
BETA-1	0.80%	4.80%	9.70%
TSMM	1.50%	4.60%	9.50%
Panel B. With contaminating events			
MANUAL	0.80%	4.50%	10.50%
BMP	1.60%	6.70%	12.10%
RANK	0.40%	3.80%	8.10%
GARCH	2.50%	6.70%	10.80%
BETA-1	0.90%	5.20%	11.00%
TSMM	0.70%	4.30%	9.80%
Panel C. Confidence intervals for rejection rate			
95% Confidence interval	0.4% 1.6%	3.7% 6.4%	8.1% 11.9%
99% Confidence interval	0.2% 1.8%	3.2% 6.8%	7.6% 12.4%

The specification analysis with no event-induced increase in return volatility, showing the rejection rates for different cross-sectional test statistics when an event creates no abnormal returns and no increase in variance. BMP corresponds to Boehmer et al.'s (1991) test, RANK is as in Corrado (1989), GARCH is the test studied by Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. TSMM and MANUAL are, respectively, the two-state market model extension and the manually filtered version of Boehmer et al.'s (1991) standardized cross-sectional approach. Panel A provides the analysis when the estimation window is not contaminated, and Panel B when it has been contaminated using the procedure described in Section 4.2. Panel C provides the confidence limits for rejection frequency in 1000 binomial trials.

For the B panels (with contaminating events) we added, in the spirit of Thompson (1988), a manually filtered approach (MANUAL). In this approach, days contaminated by simulated events during the estimation window were manually removed from the sample of observations before the implementation of the Boehmer et al. (1991) cross-sectional test of significance. This cleaning approach provides a clear benchmark against which the robustness of the set of tests used in the presence of contaminating events can be evaluated.[12]

5.1. Tests with no change in the return variance

Table 1, Panel A shows that, in the absence of contaminating events and event-induced increase in return volatility, all the tests we survey are (reasonably) well specified. We provide in Table 1 Panel C confidence intervals[13] to check whether actual rejection rates differ from the nominal ones. None of the tests present a rejection rate outside the 95% and 99% confidence

[12] It is worth noting that, using real returns, the manually filtered approach only controls for simulated events. Any real contaminating events are not neutralized by the MANUAL approach.
[13] To estimate the confidence intervals, we use the same methodology as in Cowan and Sergeant (1996).

Table 2
Rejection rates of test statistics: event-induced returns no event-induced variance

	Significance level		
	1%	5%	10%
Panel A. Without contaminating events			
BMP	52.40%	74.40%	83.50%
RANK	71.40%	88.10%	92.20%
GARCH	33.10%	52.10%	62.80%
BETA-1	29.30%	51.00%	61.40%
TSMM	56.70%	75.50%	81.90%
Panel B. With contaminating events			
MANUAL	52.30%	74.30%	81.80%
BMP	44.70%	67.40%	76.40%
RANK	65.00%	82.50%	89.00%
GARCH	26.70%	47.40%	59.90%
BETA-1	28.80%	49.60%	60.40%
TSMM	57.90%	76.30%	82.70%

The power analysis with no event-induced increase in return volatility, showing the rejection rates for different cross-sectional test statistics when an event creates an increase in returns of 1% and no increase in variance. BMP corresponds to Boehmer et al.'s (1991) test, RANK is as in Corrado (1989), GARCH is the test studied by Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. TSMM and MANUAL are, respectively, the two-state market model extension and the manually filtered version of Boehmer et al.'s (1991) standardized cross-sectional approach. Panel A provides the analysis when the estimation window is not contaminated, and Panel B when it has been contaminated using the procedure described in Section 4.2.

interval. The impact of contaminating events is analyzed in Table 1 Panel B. We get almost the same results as in Panel A. None of the rejection rates seem to be statistically different from the nominal ones.

Table 2 presents the results of the power analysis, still in the absence of event-induced increase in return volatility. Without contaminating event (Panel A), the RANK test clearly outperform the other alternatives, and the power of the BMP and the TSMM tests are quite comparable. With contaminating events (Panel B), the RANK test is still the most powerful test. However, in this setting, the TSMM test is more powerful than the BMP one, and it has a power very close to the benchmark approach, which is the MANUAL test. With and without contaminating events the GARCH and the BETA-1 are the less powerful tests. Nevertheless the results for the GARCH test must be interpreted with caution, keeping in mind that we have only simulated a one-day shock (as in, for example, Brown and Warner (1980, 1985) and Boehmer et al. (1991)) and not five consecutive shocks (as in Savickas (2003)).

Tables 1 and 2 show that the GARCH and BETA-1 approaches allow the presence of contaminating events to be controlled for. However this comes at the cost of a severe reduction in power. The RANK test appears to be the most attractive alternative: it is both less sensitive to the presence of contaminating events (in terms of specification error) and more powerful. The TSMM test is, in the absence of event-induced increase in return volatility, a second best approach.

Since in practice all events induce variance (Harrington and Shrider, in press), the most realistic case to analyze is the one where we simulate also a variance increase on the event date, and in this setting, we know from the literature that the RANK test is clearly miss-specified (see, e.g., Cowan and Sergeant, 1996; Serra, 2002; Savickas, 2003). The next sub-section is devoted to this analysis.

Table 3
Rejection rates of test statistics: no event-induced returns but event-induced variance

	Significance level		
	1%	5%	10%
Panel A. Without contaminating event			
BMP	1.40%	6.50%	11.10%
RANK	4.10%	14.40%	21.40%
GARCH	0.70%	5.60%	10.30%
BETA-1	1.10%	4.80%	10.80%
TSMM	0.80%	5.80%	10.20%
Panel B. With contaminating events			
MANUAL	1.30%	6.20%	9.50%
BMP	1.50%	5.90%	10.20%
RANK	4.40%	12.40%	20.40%
GARCH	1.10%	5.20%	11.20%
BETA-1	1.10%	3.60%	9.00%
TSMM	0.90%	4.60%	8.90%
Panel C. Confidence intervals for rejection rate			
95% Confidence interval	0.4% 1.6%	3.7% 6.4%	8.1% 11.9%
99% Confidence interval	0.2% 1.8%	3.2% 6.8%	7.6% 12.4%

The specification analysis with an event-induced increase in return volatility, showing the rejection rates for different cross-sectional test statistics when an event creates an increase in variance and no increase in returns. BMP corresponds to Boehmer et al.'s (1991) test, RANK is as in Corrado (1989), GARCH is the test studied by Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. TSMM and MANUAL are, respectively, the two-state market model version and the manually filtered version of Boehmer et al.'s (1991) standardized cross-sectional approach. Panel A provides the analysis when the estimation window is not contaminated, and Panel B when it has been contaminated using the procedure described in Section 4.2. Panel C provides the confidence limits for rejection frequency in 1000 binomial trials.

5.2. Tests with a variance increase on the event date

In Tables 3 and 4, we simulate an event-induced increase in return volatility. The results reported above for the specification analysis are qualitatively confirmed, except for the RANK test. Without any surprise, Panel A and Panel B of Table 3 indicate clearly that the RANK test is poorly specified in the presence of an event-induced increase in return volatility. This result has also been reported in previous studies.

With respect to power, Table 4 shows that the presence of an event-induced increase in return volatility strongly affects the power of all the tests (compare the rejection rates in Table 2 to those displayed in Table 4). The RANK and TSMM tests seem to be the most powerful approaches to detect the 1% simulated event-day abnormal returns. However, since these two tests differ significantly with respect to the specification error (see Table 3), it is not possible to compare their power as such. In order to overcome this problem, we resorted to a graphical method, denoted 'size–power curves', as proposed by Davidson and MacKinnon (1998).[14] These size–power curves allow the power of alternative test statistics that do not have the same size (specification) to be compared. Using the simulation techniques described in Section 4, we computed the power and size of the tests for 100 different theoretical significance levels (between 0% and 100%).

[14] Special thanks to Luc Bauwens for having suggested this method.

Table 4
Rejection rates of test statistics: event-induced returns event-induced variance

	Significance level		
	1%	5%	10%
Panel A. Without contaminating events			
BMP	18.00%	35.70%	47.50%
RANK	28.90%	48.50%	58.50%
GARCH	8.80%	21.90%	30.90%
BETA-1	8.20%	21.20%	31.70%
TSMM	18.80%	37.50%	49.80%
Panel B. With contaminating events			
MANUAL	14.20%	32.30%	44.20%
BMP	11.30%	28.40%	39.00%
RANK	21.40%	39.30%	50.10%
GARCH	5.70%	16.80%	26.60%
BETA-1	7.60%	19.70%	29.50%
TSMM	14.80%	31.90%	44.30%

The power analysis with an event-induced increase in return volatility, showing the rejection rates for different cross-sectional test statistics when an event creates an increase in returns of 1% and an increase in variance. BMP corresponds to Boehmer et al.'s (1991) test, RANK is as in Corrado (1989), GARCH is the test studied by Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. TSMM and MANUAL are, respectively, the two-state market model extension and the manually filtered version of Boehmer et al.'s (1991) standardized cross-sectional approach. Panel A provides the analysis when the estimation window is not contaminated, and Panel B when it has been contaminated using the procedure described in Section 4.2.

Fig. 2 presents the results. The size–power curve of the TSMM test clearly predominates over the curves of the alternative tests, both without and with contaminating events. However, it is important to note that the supremacy of the TSMM approach is more important when the estimation window is contaminated (Panel B). In other words, holding the size constant (or for a comparable level of size), the TSMM approach is the most powerful; BETA-1 and GARCH are the less powerful tests. Another interesting result that is worth pinpointing is that the size–power curves of the BETA-1 and GARCH tests are almost the same in Panel A (without contaminating events) and Panel B (with contaminating events), suggesting that the power of these two approaches is not significantly affected by contaminating events. This result justifies somehow the use of the BETA-1 approach in recent empirical work (e.g., in the field of M&As, see Fuller et al., 2002; Moeller et al., 2003; Aktas et al., 2006).

5.3. Practical recommendations

It follows from our analysis that the choice of method depends on the conditions of the study. As already reported in the literature (e.g., Cowan and Sergeant, 1996; Serra, 2002; Savickas, 2003), if the return variance is unlikely to increase on the event date, the RANK test offers the best specification and power for general use. However, in the presence of an event-induced increase in return volatility, the TSMM offers the best compromise between robustness to the presence of contaminating events and power.

It is important to emphasize that, according to Boehmer et al. (1991), it is not uncommon for an event to be accompanied by an increase in the cross-sectional dispersion of stock returns. More recently, Harrington and Shrider (in press) have gone a step further by stressing the fact that all

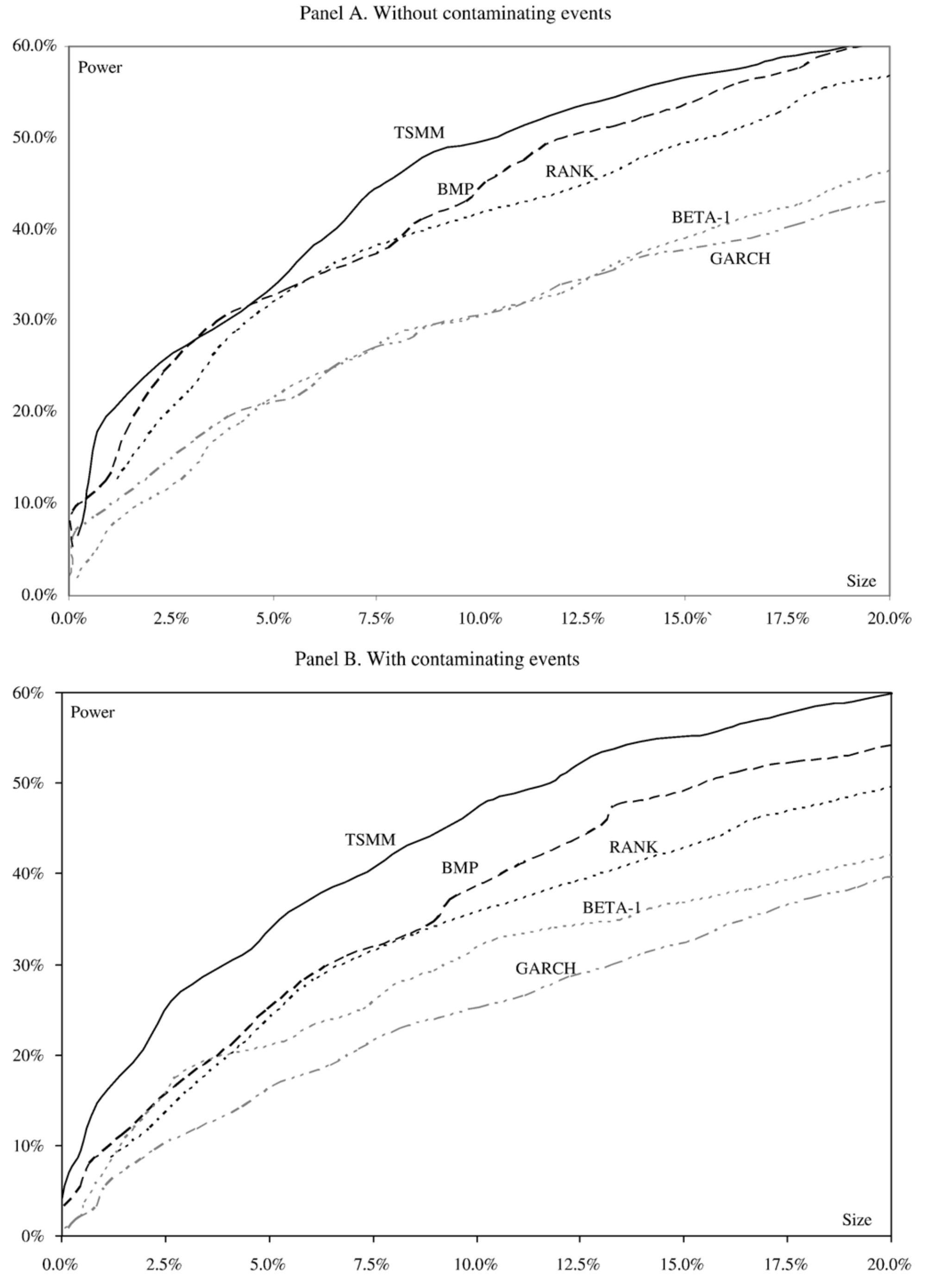

Fig. 2. This figure provides size power curves with event-induced return of 1% and an event-induced increase in return volatility. BMP corresponds to Boehmer et al.'s (1991) test, RANK is as in Corrado (1989), GARCH is the test studied by Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. TSMM is the two-state market model extension of Boehmer et al.'s (1991) standardized cross-sectional approach.

events induce variance. The key contribution of their theoretical and empirical analyses of tests for non-zero mean abnormal returns is to show that tests which are robust to cross-sectional variation should always be used. According to them the BMP approach is a good candidate for a robust test. In this paper, we showed that the TSMM test is more robust than the BMP within the context of unrelated events in the estimation window.

6. Conclusion

Analysis of the estimation window has attracted less interest in the event study literature. In this paper we have shown that unrelated events during the estimation window do affect the specification and the power of standard event-study methods. To alleviate this problem we propose the use of the TSMM approach, which is built on a two-state market model extension of Boehmer et al.'s (1991) standardized cross-sectional approach.

We have compared the TSMM approach to four alternative cross-sectional tests. These are (1) the standardized cross-sectional test (Boehmer et al., 1991), (2) the RANK test (Corrado, 1989), (3) the BETA-1 approach, which does not require any estimation window data, and (4) the GARCH-based approach introduced by Savickas (2003).

The TSMM test studied here explicitly models the volatility of the residuals as a two-state model, one corresponding to a low variance regime, and the other to a high variance regime. Our simulation results show that, in the presence of an event-induced increase in return volatility, the TSMM approach provides the best compromise in terms of the specification error and power of the test. As event-induced variance is a well-established empirical fact (see Harrington and Shrider, in press), we suggest the use of the TSMM approach.

Acknowledgements

We are grateful for constructive comments from the participants at the SIFF 2002 (Rennes, September), AFFI 2003 (Lyon, June), EFMA 2003 (Helsinki, June), and MFS 2004 (Istanbul, July) meetings. Key suggestions by Luc Bauwens, Christophe Perignon, Evangelos Sekeris, and the participants at the Paris IX-Dauphine Research Seminar (Paris, March 2003), and more specifically from Edith Ginglinger, Myron Slovin and Marie Sushka are acknowledged. We especially wish to thank Kathleen Fuller, Jeffry Netter and Mike Stegemoller who give us access to their dataset and James Hamilton for his advice. All remaining errors are ours. Part of this research was carried out while the second author was a visiting scholar in the Finance Department of the Anderson School (University of California at Los Angeles).

References

Aktas, N., de Bodt, E., Roll, R., 2004. Market response to European regulation of business combinations. Journal of Financial and Quantitative Analysis 39, 731 758.

Aktas, N., de Bodt, E., Roll, R., 2006. Hubris, learning, and M&A decisions: empirical evidence. Finance Working Paper, vol. 02 06. University of California, Los Angeles.

Asquith, P., Bruner, R.F., Mullins, D.W., 1983. The gains to bidding firms from merger. Journal of Financial Economics 11, 121 139.

Ball, C., Torous, W., 1988. Investigating security-price performance in the presence of event-date uncertainty. Journal of Financial Economics 22, 123 154.

Boehmer, E., Musumeci, J., Poulsen, A., 1991. Event study methodology under conditions of event induced variance. Journal of Financial Economics 30, 253 272.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307 327.

Brown, S., Warner, J., 1980. Measuring security price performance. Journal of Financial Economics 8, 205 258.

Brown, S., Warner, J., 1985. Using daily stock returns: the case of event studies. Journal of Financial Economics 14, 3 31.

Campbell, C., Wasley, C., 1993. Measuring security price performance using daily Nasdaq return. Journal of Financial Economics 33, 73 92.

Corrado, C., 1989. A nonparametric test for abnormal security price performance in event studies. Journal of Financial Economics 23, 385 395.

Corrado, C., Zivney, T., 1992. The specification and power of the sign test in event study hypothesis tests using daily stock returns. Journal of Financial and Quantitative Analysis 27, 465 478.

Cousin, J.G., de Launois, T., 2006. News intensity and conditional volatility on the French stock market. Finance 27, 7 60.

Cowan, A., 1992. Nonparametric event study tests. Review of Quantitative Finance and Accounting 2, 343 358.

Cowan, A., Sergeant, A., 1996. Trading frequency and event study test specification. Journal of Banking and Finance 20, 1731 1757.

Davidson, R., MacKinnon, J.G., 1998. Graphical methods for investigating the size and power of hypothesis tests. The Manchester School 66, 1 26.

Engle, R., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. Econometrica 50, 987 1008.

Fama, E., Fisher, L., Jensen, M., Roll, R., 1969. The adjustment of stock prices to new information. International Economic Review 10, 1 21.

Fuller, K., Netter, J., Stegemoller, M., 2002. What do returns to acquiring firms tell us? Evidence from firms that make many acquisitions. Journal of Finance 57, 1763 1793.

Hamilton, J., 1989. A new approach to the economic analysis of non-stationary time series and the business cycle. Econometrica 57, 357 384.

Hamilton, J., 1994. Time Series Analysis. Princeton University Press.

Harrington, S. Shrider, D. in press. All events induce variance: analyzing abnormal returns when effects vary across firms. Journal of Financial and Quantitative Analysis.

Krolzig, H.M., 1997. Markov-switching vector autoregressions. University of Oxford, UK Lecture Notes in Economics and Mathematical Systems, vol. 454. Springer. XIV.

Malatesta, P., Thompson, R., 1985. Partially anticipated events: a model of stock price reactions with an application to corporate acquisitions. Journal of Financial Economics 14, 237 250.

Mitchell, M., Stafford, E., 2000. Managerial decisions and long-term stock price performance. Journal of Business 73, 287 330.

Moeller, S., Schlingemann, F., Stulz, R., 2003. Firm size and the gains from acquisitions. Journal of Financial Economics 73, 201 228.

Nimalendran, M., 1994. Estimating the effects of information surprises and trading on stock returns using a mixed jump-diffusion model. Review of Financial Studies 7, 451 473.

Patell, J., 1976. Corporate forecasts of earnings per share and stock price behavior: empirical tests. Journal of Accounting Research 14, 246 276.

Roll, R., 1987. R-squared. Journal of Finance 43, 541 566.

Ruback, R., 1982. The effect of discretionary price control decisions on equity values. Journal of Financial Economics 10, 83 105.

Salinger, M., 1992. Standard errors in event studies. Journal of Financial and Quantitative Analysis 27, 39 53.

Savickas, R., 2003. Event-induced volatility and tests for abnormal performance. Journal of Financial Research 16, 165 178.

Schipper, K., Thompson, R., 1983. Evidence on the capitalized value of merger activity for acquiring firms. Journal of Financial Economics 11, 85 119.

Serra, A.P., 2002. Event study tests: a brief survey. Working Paper FEP, vol. 117. University of Porto.

Sharpe, W., 1963. A simplified model for portfolio analysis. Management Science 9, 277 293.

Thompson, J., 1988. More methods that make little difference in event studies. Journal of Business Finance and Accounting 15, 77 86.

Exhibit D

Thompson (1988)

Journal of Business Finance & Accounting, 15(1) Spring 1988, 0306 686X $2.50

MORE METHODS THAT MAKE LITTLE DIFFERENCE IN EVENT STUDIES

JOEL E. THOMPSON*

Security return event studies are one of the most objective ways in which researchers can determine whether accounting or other events have an impact on investors. The viability of such research depends upon the ability of the methods employed to detect changes in security returns that may be caused by the event of interest.

Many methods are available for detecting changes in security returns (see Beaver, 1982). The choice among them has been difficult and has often resulted in the researcher conducting numerous redundant tests. However, it has been recently shown by Brown and Warner (1980 and 1985) and Dyckman et al. (1984) that many event study methods produce about the same number of Type I errors and power.

The purpose of this study is to investigate three additional areas to determine their importance in event studies. In particular, industry indexes, return form, and extraneous individual firm events are considered. Using the simulation technique introduced by Brown and Warner (1980), it is found that each of these factors make little difference in event studies. The usual market model approach with simple returns and with ignoring extraneous individual firm events should be sufficient in most cases. Thus, this research should help to further simplify the choice of event study methods.

THE METHODS

Industry factors have been taken into account in several event studies (e.g., Smith, 1981; and Bell, 1983) and they have been considered by Beaver (1981). In addition, the relationship of industry factors to firm stock returns has been investigated by King (1966), Meyers (1973), Fertuck (1975), and others. However, these studies do not directly and empirically determine the importance of industry factors in tests for the effects of events on security returns. Hence, in this study the market, industry, and market-industry models are

* The author is Assistant Professor on the Faculty of Accounting and Management Information Systems at The Ohio State University. He would like to thank the participants of the Finance Seminar and the Accounting Research Colloquium both at The Ohio State University, Department of Accounting Seminar at Kent State University, Bruce Gaumnitz, Joanne Turner, David Williams, and an anonymous referee for comments on earlier versions of this paper. He is also grateful to his Faculty and College who partially supported this research through a grant. (Paper received January 1986, revised July 1986)

empirically compared and are, respectively:

$$R_{it} = \alpha_{i0} + \alpha_{i1}R_{mt} + \epsilon_{it} \quad (1)$$

$$R_{it} = \beta_{i0} + \beta_{i1}R_{It} + \delta_{it} \quad (2)$$

$$R_{it} = \lambda_{i0} + \lambda_{i1}R_{mt} + \lambda_{i2}R_{It} + \theta_{it} \quad (3)$$

where: R_{it} is the simple return of firm i in period t; R_{mt} is a market index; R_{It} is an industry index; α_{i0}, α_{i1}, β_{i0}, β_{i1}, λ_{i0}, λ_{i1}, and λ_{i2} are coefficients; and ϵ_{it}, δ_{it}, and θ_{it} are the error terms for firm i in period t with the usual assumptions of independence, normality, zero mean, and constant variance. All of the models can be derived under the assumption that the joint distribution of returns is multivariate normal.[1] However, it is an empirical question as to the relative numbers of Type I errors and powers that result from using these models in statistical tests. Type I errors and powers will depend on such things as the strength of the relationship between R_{it} and the indexes and the appropriateness of the normality assumption.

Another empirical question confronting a researcher is whether to use simple or continuously compounded returns (e.g., see Beaver, 1982).[2] Continuously compounded returns are more consistent with a normal distribution than simple returns under certain assumptions (see Fama, 1976, pp. 17–20). Hence, the actual proportion of Type I errors in a large number of tests when using continuously compounded returns may be in closer agreement with the preset significance levels. In addition, since the natural logarithm transformation used to obtain continuously compounded returns reduces positive simple returns and increases in absolute value negative simple returns, positive event effects should be easier to detect using simple returns while negative effects should be easier to detect using continuously compounds returns.[3] And more dramatic differences should be evident in left tail tests given that the transformation has a larger impact on negative simple returns than on positive simple returns. However, in each of these comparisons, the degree of difference is not readily determinable analytically.

The last issue examined in this study deals with the importance of extraneous individual firm events. Researchers are concerned with the occurrence of other events contemporaneously with the event being studied. Such extraneous events can result in rejecting the null hypothesis when it is true (i.e., a Type I error). In this study, the impact of extraneous individual firm events on Type I errors is examined. In addition, the importance of extraneous individual firm events occurring during the estimation period is investigated. These extraneous events may increase the variance of a firm's returns resulting in a larger estimated variance and thereby decrease the power of the test. Hence, returns associated by time with such events during the estimation period are deleted to determine whether this procedure results in a more powerful test.

PROCEDURES

The methods are compared using a simulation technique similar to that of Brown and Warner (1980 and 1985) and Dyckman et al. (1984). That is, hypothetical events are assumed to take place in randomly selected time periods. Then, various hypothetical effects are added to the firm returns during the hypothetical event periods to determine how well the effects can be detected in tests based on the methods.

The methods are compared for Type I errors and power in individual firm t-tests which are consistent with a normality assumption for the error terms in the models and which have been used in several event studies (e.g., Lev, 1979; Gheyara and Boatsman, 1980; and Bell, 1983). The null hypothesis of the individual firm t-test is that the effect of an event of interest on a firm's return during the event period is 0. The general form of the test statistic is

$$(R_{i0} - P_{i0})/SE_{i0} \tag{4}$$

which has a t-distribution with n-p degrees of freedom when the null hypothesis is true where: period 0 is the event period; P_{i0} is the prediction of firm's i return in event period 0; SE_{i0} is the estimated standard error (see Neter and Wasserman, 1974, p. 233); n is the number of observations used to estimate the model; and p is the number of coefficients in the model (i.e., see equations (1–3)).[4]

The simulation results are based on eleven two-digit SIC industries which were randomly selected without replacement from the *Standard Industrial Classification Manual*. These industries contained 465 American and New York Stock Exchange firms which had complete return data on the Center for Research in Security Prices (CRSP) computer tapes for the 66 consecutive trading day period randomly selected for each industry.[5] The first 60 days were used as the estimation period while the last six days were treated as event periods.[6] The industry index for a given firm was the equally weighted average of the other firms' returns in the same industry. The three smallest industry indexes consisted of the returns of 3, 10, and 16 firms while the three largest industry indexes consisted of the returns of 69, 95, and 97 firms. The equally weighted market index from the CRSP tape was used.[7] The models were estimated using ordinary least squares.

To compare Type I errors and power, six samples of 465 individual firm t-tests were constructed. The first sample is comprised of the tests for the first of six event days. The other samples are defined analogously. The proportion of Type I errors for the t-tests for a given method is the proportion of rejections of the null hypothesis when no hypothetical effect is added to the firm returns. A preset significance level of 0.025 was used in both left and right tail tests. To compare powers, the numbers ±0.01 and ±0.05 were added

proportion of Type I errors for continuously compounded returns are closer to the preset significance level than the proportion of Type I errors for simple returns. In addition, the differences between panels are greater for left tail tests than right tail tests. These results are consistent with the nature of the transformation from simple to continuously compounded returns discussed previously. But only marginal improvement in power is realized by changing the return form and this improvement in power is accompanied by a slight increase in Type I errors. Thus, return form also does not seem to be an important consideration in event studies.

It is interesting to note that the distribution of the prediction errors for each of the models appears assymmetrical as evidenced by the proportions of Type I errors. Under certain assumptions, the 95 percent confidence interval for the proportion of Type I errors is 0.0192 to 0.0408 (Mood et al., 1974, pp. 394–395). Thus, the proportions of Type I errors differ significantly from the preset significance level of 0.025 in right tail tests but not in left tail tests. This asymmetry and apparent non-normality is explored by examining the empirical distribution of the returns themselves during the event periods. Table 2 shows that both the simple and continuously compounded return distributions seem asymmetrical and non-normal. Their respective chi-square goodness of fit statistics (Conover, 1971, pp. 186–187), based on the categories shown in Table 2, of 254.32 and 243.97 imply that it is highly unlikely (significance level <0.001) that these returns are from a normal distribution. Moreover, these distributions have the characteristic leptokurtic shape reported by Fama (1976, pp. 21–24).

This non-normality may be due to the non-synchronous trading problem discussed by Scholes and Williams (1977). However, it does not appear that ordinary least squares, though it may result in biased coefficients in the models for firms whose stock trades very infrequently or very frequently, significantly distorts the distributions of the prediction errors of the models relative to that of the actual returns themselves.[11] In particular, note that the proportions of observations in Table 2 falling beyond two standard errors (which corresponds with a preset significance level of 0.025) is very similar to the proportions of Type I errors reported in Table 1. This lack of significant distortion by ordinary least squares may partly explain why Scholes and Williams' (1977) or Dimson's (1979) estimation procedures do not appear to affect the proportion of Type I errors or power in event studies (Dyckman et al., 1984; and Brown and Warner, 1985).

Lastly, the importance of extraneous individual firm events is examined by using firm events cited in the *Wall Street Journal Index (WSJI)*. Individual firm returns occurring the trading day before, the trading day of, or the trading day after a firm event cited in the *WSJI* were deleted from that firm's estimation period.[12] This was done for each of the 332 firms which did not have such individual firm events occurring during the event period or within one day of them. The proportion of Type I errors and powers are reported in Table

Table 2

Comparison of the Empirical Distribution of the Returns During the Event Periods (2,790 observations) with that of a Normal Variable.

*Range of a Standardized Variable X**	*Frequency (%) for a Normal Variable*	*Simple Returns*		*Compound Returns*	
		Actual Freq. (%)	*Deviation from Normal*	*Actual Freq. (%)*	*Deviation from Normal*
$X \le -3.0$	0.2	0.7	0.5	0.8	0.6
$-3.0 < X \le -2.5$	0.6	0.5	−0.1	0.4	−0.2
$-2.5 < X \le -2.0$	1.7	1.0	−0.7	1.0	−0.7
$-2.0 < X \le -1.5$	4.4	3.2	−1.2	3.2	−1.2
$-1.5 < X \le -1.0$	9.1	8.0	−1.1	8.0	−1.1
$-1.0 < X \le -0.5$	14.9	14.7	−0.2	14.4	−0.5
$-0.5 < X < 0$	19.1	26.6	7.5	26.0	6.9
$0 \le X < 0.5$	19.1	18.1	−1.0	18.9	−0.2
$0.5 \le X < 1.0$	14.9	13.2	−1.7	13.2	−1.7
$1.0 \le X < 1.5$	9.1	7.1	−2.0	7.4	−1.7
$1.5 \le X < 2.0$	4.4	3.5	−0.9	3.4	−1.0
$2.0 \le X < 2.5$	1.7	1.6	−0.1	1.8	0.1
$2.5 \le X < 3.0$	0.6	0.8	0.2	0.6	0.0
$3.0 \le X$	0.2	1.0	0.8	0.9	0.7

* The empirical returns for each firm were standardized using estimates of the mean and standard error based on each firm's respective estimation period. Thus, if the returns are actually normally distributed, the standardized empirical returns should have a t-distribution with 59 degrees of freedom. This t-distribution is reported in the second column of the table.

Table 3

Average Proportions of Rejections of the Null Hypothesis of One-tail Individual Firm t-tests in Samples of 332 Firms Using Simple Returns with (Partial Estimation Period) and Without (Full Estimation Period) Eliminating Firm Returns Associated by Time with Extraneous Events.

Six samples (i.e. event days) were used and the preset significance level was 0.025.

	Left Tail Tests			*Right Tail Tests*		
Hypothetical Effects	0	−0.01	−0.05	0	0.01	0.05
Model:						
Panel A — Partial Estimation Period:						
Market	0.0200	0.0648	0.6290	0.0382	0.0838	0.6511
Industry	0.0226	0.0618	0.6461	0.0377	0.0808	0.6496
Market-Industry	0.0211	0.0633	0.6466	0.0377	0.0818	0.6536
Panel B — Full Estimation Period:						
Market	0.0181	0.0597	0.6170	0.0367	0.0783	0.6295
Industry	0.0206	0.0572	0.6325	0.0331	0.0728	0.6290
Market-Industry	0.0211	0.0577	0.6355	0.0356	0.0758	0.6340

3 along with the results using a full estimation period. Examination of this table reveals that power can be improved by using such a partial estimation period, but again the improvement is small and tends to be accompanied by an increase in Type I errors.[13]

With respect to the importance of individual firm extraneous events in or near the event period, note that the proportion of Type I errors in Panel A of Table 3 are very close to those reported in Panel A of Table 1. In fact, this is expected since extraneous events in the event period (for Table 1 results) should be controlled to the extent that they and the extraneous events in the (full) estimation period produce similar effects on a firm's returns. In other words, the preset significance level for Table 1 results already takes into account typical individual firm extraneous events.

Thus, one implication for event studies is that extraneous individual firm events occurring during the estimation period have little impact on the power of the tests. The other implication is that extraneous individual events occurring during the event period have little impact on Type I errors to the extent that they produce effects on returns similar to those which result from extraneous events occurring during the estimation period. Of course, atypical extraneous events occurring during or near an event period can have an effect on Type I errors and must be of concern to researchers. Exactly what constitutes an atypical event remains an issue for future research.

LIMITATIONS AND CONCLUDING REMARKS

The results are based on individual firm tests. The validity of these results for a sample of firms tested would seem to depend on cross-sectional dependence. If the prediction errors of each of the models during the event period are cross-sectionally independent, then the comparisons of the methods should be essentially the same using most sample statistical tests. Power in this case should be determined by the amount of noise in the individual firm prediction errors (i.e., their standard errors).

The comparisons of the methods in sample tests when there is cross-sectional dependence is not as clear. However, it is interesting to note that the market-industry model performed about the same as the market model in this study. Thus, industry must not be a strong source of cross-sectional dependence as is typically thought.

In any case, with respect to at least individual firm t-tests, it was found in this study that industry indexes, return form, and individual firm extraneous events have little impact on event study results. Perhaps refinements such as examining particular industries or types of individual firm events would identify instances where these factors made a significant difference. But in general, the factors examined in this study do not appear to have an important role in event studies.

NOTES

1 A derivation of the market model is given by Fama (1976, pp. 66–72) and an analogous argument can be used to derive the industry model. By extending these arguments to three variables (based on Theorem 2.5.1 of Anderson, 1958, p. 29) the market-industry model is derived (Fama, 1976, p. 371). Note that R_{mt} and R_{It} do not need to be independent.

2 Brown and Warner (1985) do examine the issue of simple versus continuously compounded returns. However, they merely state that the results are similar and do not give any explicit evidence.

3 The continuously compounded return is $\ln(1 + R_{it})$ where ln is the natural logarithm function.

4 It is assumed that the variance of a firm's returns remains the same from the estimation to the event period. Several researchers have been concerned that the variance of a firm's return in the event period may be larger than the variance in the estimation period (see Collins and Dent, 1984). But, as a practical matter, whether the impact of the event of interest on a firm's return is through a mean effect or variance effect on its distribution makes little difference in the application of the individual firm t-tests whose statistic is derived under the null hypothesis of no event effect (whatsoever) on the firm's return. Either a mean or variance effect can make a firm's return unusually large and the null hypothesis will be rejected.

5 To be included an industry required at least two firms with complete data so that an industry index could be constructed. Also, the last day of the 66 day period was randomly chosen from the period January 1, 1976 through December 31, 1980 which is a relatively recent time period in which researchers may wish to apply some of the methods discussed in this paper.

6 Using six event days per firm is an efficient method of expanding the sample size (a similar procedure was used by Dyckman et al., 1984). Thus, a total of 6 × 465 = 2,790 individual firm t-tests are conducted for each method which should be sufficient to distinguish any important differences in Type I errors and power. For example, under certain assumptions (see Mood et al., 1974, pp. 394–395), 2,790 observations implies at a 95 percent confidence level that

the proportion of Type I errors should be within ± 0.006 of a preset significance level of 0.025.

7 Brown and Warner (1985) found that a value weighted market index produced similar results with respect to the market model.

8 No hypothetical effects were added to the market or industry indexes.

9 There should be some cross-sectional dependence in the prediction errors of the firms in the same industry since the same event periods are used for these firms. But cross-sectional dependence will not affect the average proportions of rejections over time that are reported in this study. That is, the average proportions of rejections are still unbiased estimates of the true proportions of rejections.

10 The individual firm simple returns were transformed to continuously compounded returns before computing the industry index. The equally weighted market index on a continuously compounded basis was approximated as $\ln(1 + R_{mt})$ where R_{mt} is the equally weighted market index using simple returns. Also, the hypothetical effects were added to the simple returns before transformation to continuously compounded returns.

11 Under the assumptions used by Scholes and Williams (1977, p. 313), the means used to standardize the returns in this study are unbiased estimates of the means of the unobservable true returns.

12 Occasionally, the *WSJ* is published on non-trading days. In these cases, the trading day of occurrence was defined as the first trading day after the *WSJI* citation date.

13 This difference in Type I errors can be explained by the sample, which was restricted to firms which should have had larger returns in absolute value during the full estimation period compared to the event period, since all but four of these firms had *WSJI* citations during the full estimation period but none of the firms had citations during the event periods or within one day of it. Thus, the smaller proportion of Type I errors is expected when using a full estimation period for these firms.

REFERENCES

Anderson, T. (1958), *An Introduction to Multivariate Statistical Analysis*, (John Wiley & Sons, 1958).

Beaver, W. (1981), 'Econometric Properties of Alternative Security Return Methods', *Journal of Accounting Research* (Spring 1981), pp. 163–184.

______ (1982), 'Discussion of Market-Based Empirical Research in Accounting: A Review, Interpretation and Extension', *Journal of Accounting Research* (Supplement 1982), pp. 323–331.

Bell, T. (1983), 'Market Reaction to Reserve Recognition Accounting', *Journal of Accounting Research* (Spring 1983), pp. 1–17.

Brown, S. and J. Warner (1980), 'Measuring Security Price Performance', *Journal of Financial Economics* (September 1980), pp. 205–258.

______ (1985), 'Using Daily Stock Returns: The Case of Event Studies', *Journal of Financial Economics* (March 1985), pp. 3–31.

Collins, D. and W. Dent (1984), 'A Comparison of Alternative Testing Methodologies Used in Capital Market Research', *Journal of Accounting Research* (Spring 1984), pp. 48–84.

Conover, W. (1971), *Practical Nonparametric Statistics* (John Wiley & Sons, 1971).

Dimson, E. (1979), 'Risk Measurement When Shares Are Subject to Infrequent Trading', *Journal of Financial Economics* (June 1979), pp. 197–226.

Dyckman, T., D. Philbrick, and J. Stephan (1984), 'A Comparison of Event Study Methodologies Using Daily Stock Returns: A Simulation Approach', *Journal of Accounting Research* (Supplement 1984), pp. 1–30.

Fama, E. (1976), *Foundations of Finance* (Basic Books, 1976).

Fertuck, L. (1975), 'A Test of Industry Indices Based on SIC Codes', *Journal of Financial and Quantitative Analysis* (December 1975), pp. 837–848.

Gheyara, K. and J. Boatsman (1980), 'Market Reaction to the 1976 Replacement Cost Disclosures', *Journal of Accounting and Economics* (August 1980), pp. 107–125.

King, B. (1966), 'Market and Industry Factors in Stock Price Behavior', *Journal of Business* (January 1966), pp. 139–190.

Lev, B. (1979), 'The Impact of Accounting Regulation on the Stock Market: The Case of Oil and Gas Companies', *The Accounting Review* (July 1979), pp. 485–503.

Meyers, S. (1973), 'A Re-examination of Market and Industry Factors in Stock Price Behavior', *Journal of Finance* (June 1973), pp. 695–705.

Mood, A., F. Graybill and D. Boes (1974), *Introduction to the Theory of Statistics*, third edition (McGraw-Hill, 1974).

Neter, J. and W. Wasserman (1974), *Applied Linear Statistical Models* (Richard D. Irwin, 1974).

Scholes, M. and J. Williams (1977), 'Estimating Betas from Non-synchronous Data', *Journal of Financial Economics* (December 1977), pp. 309–327.

Smith, A. (1981), 'The SEC "Reversal" of FASB Statement No. 19: An Investigation of Information Effects', *Journal of Accounting Research* (Supplement 1981), pp. 174–211.

Statistical Policy Division, Executive Office of the President-Office of Management and Budget (1972), *Standard Industrial Classification Manual* (U.S. Government Printing Office, 1972).